Stat & Data Analysis: Final Exam Review NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

MULTIPLE-CHOICE: The questions or incomplete statements that follow are each followed by five suggested answers or completions. Choose the response that best answers the question or completes the statement. CIRCLE THE LETTER.

***CHAPTER 1***

19. A measurement process that does not systematically overstate or understate the true value of the quantity being measured is called

(a) valid. (b) reliable. (c) random. (d) unbiased.

25. The net asset value of a mutual fund has increased from $27 on December 31 to $33 now. The percent increase in value is about

(a) 22%. (b) 18%. (c) 1.2%. (d) 122%. (e) 82%

27. Michelle's income two years ago was $420,000. Last year her income was only $100,000. The percent change in Michelle's income was

(a) 76.2% decrease (b) 320% decrease (c) 31.3% decrease (d) 23.8% decrease

(e) None of these

28. The average wage of production workers (adjusted for the effects of inflation) was $11.08 an hour in 1981 and $10.35 an hour in 1991. In the decade of the 1980s, wages went down by about

(a) 73%. (b) 7.3%. (c) 7.0%. (d) 6.6%.

30. A researcher creates a machine that will measure the total electrical activity in a human brain over a short period of time. She claims that this measures the intelligence of the brain's owner. A psychologist says that's not true because the amount of electrical activity isn't related to ability to solve problems. The psychologist is claiming that electrical activity as a measure of intelligence is

(a) invalid. (b) biased. (c) not reliable. (d) not accurate.

32. A local police department gives everyone who applies for a job a test in American history.

However, experience shows that these test scores are unrelated to future job performance.

As a measure of ability to do police work, the history test scores

(a) are not reliable.

(b) are biased.

(c) are confounded.

(d) are invalid.

(e) have predictive validity.

33. The same police department also gives job applicants a test of their knowledge of modern popular music. Experience shows that those who score well on this test tend to become lousy police officers. As a measure of future job performance, the music test scores

(a) are not reliable.

(b) are biased.

(c) are confounded.

(d) are invalid.

(e) have predictive validity.

34. In 1970, 8.4 billion cigars were smoked in the United States. In 1989, 2.0 billion cigars were smoked. What percentage change is this?

(a) 7.6% decrease

(b) 7.6% increase

(c) 176% decrease

(d) 76% increase

(e) 76% decrease

36. A dishonest butcher has a scale on which he weighs the meat his customers buy. In order to increase his profits, he has doctored the scale so that it always reads very close to 10 percent more than the actual weight. The measurements from this scale are

(a) biased and unreliable

(b) biased and reliable

(c) unbiased and unreliable

(d) unbiased and reliable

37. An ad from a local appliance store says, "Double the Difference Price Protection: If, during the first 30 days from the date you purchase a product from H. H. Gregg, you find the same item at a lower price at another store we will refund 200% of the difference." What does this mean?

(a) It means that H. H. Gregg will reduce its price by 200%.

(b) It means that H. H. Gregg will reduce its price to one-third of what it was.

(c) It makes perfectly good sense, as long as the other store's price is at least half of H. H. Gregg's price.

(d) It's nonsense, because refunding 100 percent of the difference already reduces the cost to zero.

(e) It's nonsense because percents only make sense for counts, and the price of an appliance isn't a count.

38. A student doing a science fair project tries to germinate tomato seeds at different soil temperatures. She writes, "I planted 10 seeds at each of three temperatures. I found that

20% germinated at 55, 40% germinated at 60, and 37% germinated at 65." Why must

her report be wrong?

(a) 37% is not a possible percent in this situation.

(b) The three percents given don't add to 100%.

(c) It's wrong to report percents; she should report the correlation *r*.

(d) This isn't a randomized comparative experiment.

(e) It isn't possible for fewer seeds to germinate at 60than at 65

***CHAPTER 2***

A recent Gallup poll asked "Do you consider pro wrestling to be a sport, or not?" Of the people asked, 81% said "No." Here is what Gallup says about the accuracy of this poll:

The results below are based on telephone interviews with a randomly selected national sample of 1,028 adults, 18 years and older, conducted August 16-18, 1999. For results based on this sample, one can say with 95 percent confidence that the maximum error attributable to sampling and other random effects is plus or minus 3 percentage points. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.

**The next three questions** concern this situation.

14. The population for this poll appears to be

(a) all adults, 18 years and older.

(b) 95% of adults, 18 years and older.

(c) the 1028 adults who were interviewed.

(d) 95% of the 1028 adults who were interviewed.

15. The sample for this poll is

(a) all adults, 18 years and older.

(b) 95% of adults, 18 years and older.

(c) the 1028 adults who were interviewed.

(d) 95% of the 1028 adults who were interviewed.

18. Which of these sources of possible errors in the poll result are examples of nonsampling errors?

(a) The poll left out people without telephones.

(b) Some people chosen for the sample refused to answer.

(c) Some people did not tell the truth because they were embarrassed to admit they like pro wrestling.

(d) None of these.

(e) (b) and (c) but not (a).

28. When you drop your pencil point blindly into the middle of a table of random digits, what is the chance that the three digits to the right of where you land will be 999?

(a) 1 in 100, because every three-digit group has the same chance to come up.

(b) 1 in 1000, because every three-digit group has the same chance to come up.

(c) no chance, because 999 is not a random group of digits.

(d) can't say -- it is completely random.

32. When we take a census, we attempt to collect data from

(a) a stratified random sample

(b) every individual selected in a simple random sample

(c) every individual in the population

(d) a voluntary response sample

(e) a convenience sample

A national sample survey interviewed 3,800 people age 18 and older nationwide by telephone. One question asked was whether they agreed with this statement: "Some people say we should have a third major political party in this country in addition to the Democrats and Republicans." **The next three questions** concern this opinion poll

52. The population for this sample survey appears to be

(a) all adult residents of the U.S.

(b) all registered voters.

(c) the 3,800 people who were interviewed.

(d) people who think we should have a third political party.

(e) all U.S. citizens.

53) In the survey, 53% of the people asked agreed that we should have a third party. The number 53% is a

(a) correlation. (b) parameter. (c) margin of error. (d) confidence level. (e) statistic.

56. A possible source of nonsampling error in this sample survey is

(a) some people chosen for the sample refused to answer questions.

(b) people without telephones could not be in the sample.

(c) some people never answered the phone in several calls.

(d) Both (a) and (c).

(e) All of (a), (b), and (c).

59. You want to take an SRS of 50 of the 816 students who live in a college dormitory. You label the students 001 to 816 in alphabetical order. In the table of random digits you read the entries

96746 12149 37823 71868 18442 35119 62103 39244

The first three students in your sample have labels

(a) 967, 461, 214 (b) 967, 121, 378 (c) 461, 214, 937

(d) 461, 214, 718 (e) 674, 612, 149

61. Which of these statements about the table of random digits is true?

(a) Every row must have exactly the same number of 0's and 1's.

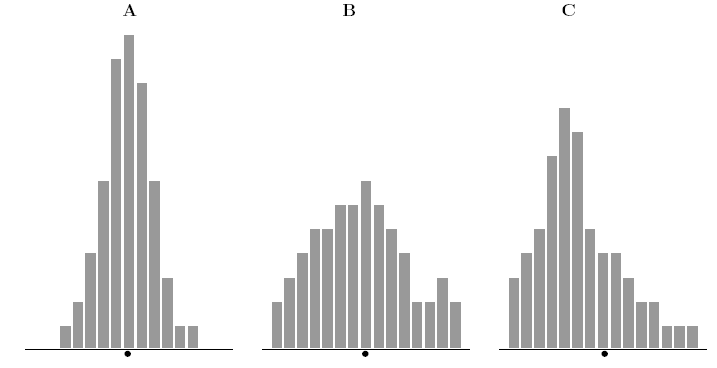
(b) In the entire table, there are exactly the same number of 0's and 1's.

(c) If you look at 100 consecutive pairs of digits anywhere in the table, exactly 1 pair is 00.

(d) All of these are true.

(e) None of these is true.

The histograms below describe the values taken by three sample statistics in several hundred samples from the same population. The true value of the population parameter is marked on each histogram. **The next four questions** refer to these histograms.



70. The name for the pattern of values that a statistic takes when we sample repeatedly from the same population is

(a) the bias of the statistic.

(b) the distribution of the statistic.

(c) the scale of measurement of the statistic.

(d) the variability of the statistic.

(e) the sampling error.

71. The statistic that has the largest bias among these three is

(a) A. (b) B. (c) C.

(d) A and B have similar bias, and it is larger than the bias of C.

(e) B and C have similar bias, and it is larger than the bias of A.

72. The statistic that has the lowest variability among these three is

(a) A. (b) B. (c) C.

(d) A and B have similar variability, and it is less than the variability of C.

(e) B and C have similar variability, and it is greater than the variability of A.

73. Based on the performance of the three statistics in many samples, which is preferred as an estimate of the parameter?

(a) A. (b) B. (c) C.

(d) either A or B would be equally good.

(e) either B or C would be equally good.

74. You plan to give a math achievement test to samples of 15 year-olds from both the U.S. and Korea in order to compare mathematics knowledge in the two countries. In each country, you will choose

300 students from low-income families

400 students from middle-income families

200 students from high-income families

The sample from Korea is a

(a) multistage sample.

(b) simple random sample.

(c) convenience sample.

(d) voluntary response sample.

(e) stratified sample.

75. You want to know the opinions of American school teachers about establishing a national test for high school graduation. You obtain a list of the members of the National Education Association (the largest teachers' union) and mail a questionnaire to 2500 teachers chosen at random from this list. In all 1347 teachers return the questionnaire. In this situation, the *population* is

(a) the 1347 teachers who mail back the questionnaire.

(b) the 2500 teachers to whom you mailed the questionnaire.

(c) all members of the National Education Association.

(d) all American school teachers.

(e) all American school students.

77. In the situation of the two previous questions, the *sample* is

(a) the 1347 teachers who mail back the questionnaire.

(b) the 2500 teachers to whom you mailed the questionnaire.

(c) all members of the National Education Association.

(d) all American school teachers.

(e) all American school students.

82. An example of a nonsampling error that can reduce the accuracy of a sample survey is:

(a) Using voluntary response to choose the sample.

(b) Using the telephone directory as the sampling frame.

(c) Interviewing people at shopping malls to obtain a sample.

(d) Variation due to chance in choosing a sample at random.

(e) Many members of the sample cannot be contacted.

83. In a table of random digits it is true that

(a) every pair of digits has chance 1/100 to be any of the 100 possible pairs 00, 01,…, 99.

(b) if a pair of digits is 00, the next pair cannot also be 00.

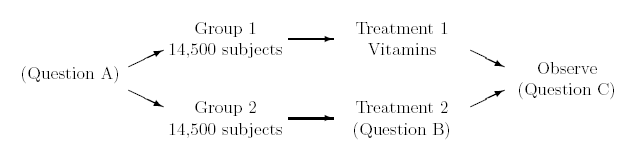
(c) every row has exactly the same number of 0's and 1's.

(d) (a) and (b) but not (c).

(e) All of (a), (b), and (c).

***CHAPTER 3***

Does taking large amounts of vitamins protect against cancer? To study this question, researchers enrolled 29,000 Finnish men, all smokers over the age of 50. Half of the men took vitamin supplements, and others took a dummy pill that has no active ingredient. The researchers followed all the men for eight years. At the end of the study, men in the vitamin group were no less likely to have cancer than men in the other group. This study cast doubt on the popular idea that taking lots of vitamins can reduce the risk of cancer. **The next six questions** are based on this study. The study design looked like this:



46. The statistical name for this study design is

(a) simple random sample.

(b) stratified random sample.

(c) randomized comparative experiment.

(d) multistage sample.

(e) observational study.

47. The method used to form the groups should appear in the outline at the point marked (Question A). What is this method?

(a) Random allocation.

(b) Voluntary response.

(c) First come, first served.

(d) The double-blind method.

(e) Stratified sampling.

48. Treatment 2 was a dummy pill. Such a dummy treatment is called a

(a) double blind.

(b) categorical variable.

(c) nonsampling error.

(d) placebo.

(e) comparative.

49. The response variable should be named in the outline at (Question C). The response variable in this study is

(a) whether or not a subject took vitamins.

(b) 29,000 Finnish men.

(c) random allocation.

(d) a confidence statement.

(e) whether or not a subject developed cancer.

50. In order to avoid unconscious bias, neither the subjects not the doctors who examined them knew whether a particular subject was taking vitamins or dummy pills. This is called

(a) the placebo effect.

(b) the double-blind technique.

(c) the retrospective method.

(d) stratified sampling.

(e) internal validity.

51. A weakness of this study is that

(a) it isn't clear that the results apply to women.

(b) observational studies give only weak evidence for causation.

(c) the people who took vitamins may have had other good habits.

(d) the response is measured in a biased way.

(e) nonsampling errors may be large.

**The next two questions** are based on the following excerpt from a newspaper article. Women who are obese suffer social and economic consequences, a new study has shown. They are 20 percent less likely to marry, have household incomes that average of $6,710 lower and are 10 percent more likely to live in poverty. The findings are from an study of 10,039 randomly selected people who were 16 to 24 years old when the research began.

62. This study was based on

(a) a randomized comparative experiment.

(b) a matched-pairs experiment.

(c) a voluntary response sample.

(d) a probability sample.

63. Does this study give strong evidence that being obese *causes* lower income?

(a) Yes, because the study included both people who were fat and people who were not.

(b) Yes, because the subjects in the study were selected at random.

(c) No, because the study showed that there is no connection between income and being fat.

(d) No, because people who are already poor may tend to overeat and/or get less exercise, which may make them more likely to be fat.

74. A recent medical study found that people who drink more than 4 cups of coffee a day have more heart attacks than people who drink less coffee or no coffee. This led some doctors to suspect that coffee may be a contributing factor in causing heart attacks. However, more careful analysis of the data showed that heavy coffee drinkers tend to smoke more than other people. This is an example of

(a) the placebo effect

(b) response error

(c) the double-blind technique

(d) a completely randomized design

(e) confounding

76. The Physicians' Health Study gave half of a group of 22,000 male M.D.'s aspirin. The other half received a placebo. After five years there were 189 heart attacks in the placebo group and 104 in the aspirin group. One of the response variables in the Physicians' Health Study was

(a) the 22,000 male M.D.'s.

(b) whether a subject took aspirin or a placebo.

(c) whether or not a subject had a hear attack.

(d) the 11,000 subjects who received aspirin.

77. A psychologist recently said that "For relatively mild medical problems, the placebo effect will produce positive results in roughly two-thirds of patients." The placebo effect is

(a) the bias due to voluntary response in a sample.

(b) the effect of a dummy treatment on a patient.

(c) a violation of confidentiality.

(d) the effect of confounding in an observational study.

A study at Yale Medical School investigated treatments for chronic cocaine abusers. Of the 72 subjects, all cocaine users who wanted to quit, 24 were assigned at random to each of three drugs: (a) desipramine, (b) lithium, (c) a placebo. After 6 weeks the subjects reported whether they had stayed off cocaine. **The next four questions** concern this study.

78. The Yale study is an example of

(a) an experiment.

(b) a sample survey.

(c) using available data.

(d) an observational study, but not a sample survey.

79. The explanatory variable in the Yale study is

(a) whether or not a subject stayed off cocaine. (b) 72 cocaine users.

(c) which drug a subject was given. (d) randomization.

80. The response variable in the Yale study is

(a) whether or not a subject stayed off cocaine. (b) 72 cocaine users.

(c) which drug a subject was given. (d) randomization.

81. All of the subjects in the Yale study were male. A similar study had 60 male and 30 female subjects. The men and women were separately assigned at random to the three treatments so that separate conclusions are possible for men and women. This design is called

(a) a block design.

(b) a stratified random sample.

(c) a matched-pairs design.

(d) a completely randomized design.

***CHAPTER 4***

79. The mean and the median of a distribution are measures of

(a) the number of observations

(b) association

(c) variability or spread

(d) center

The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380

students took the test. **The next three questions** concern this distribution.

83. About how many students had scores between 35 and 61?

(a) 35

(b) 61

(c) 95

(d) 190

(e) 285

90. The five-number summary of the distribution of scores on a statistics exam is

0 26 31 36 50

316 students took the exam. About how many students had scores above 26?

(a) 298

(b) 287

(c) 237

(d) 212

(e) can't be determined without more information.

**The next two questions refer to the following information**: The five-number summary of the distribution of scores on the final exam in Psych 001 last semester was

18 39 62 76 100

A total of 416 students took the exam.

94. About how many students had scores above 39?

(a) 416

(b) 312

(c) 104

(d) 400

95. The 80th percentile was

(a) 76

(b) between 62 and 76

(c) between 76 and 100

(d) probably between 39 and 76, since most of the class scored between these two numbers.

***CHAPTER 5***

2. Were the extinctions that occurred in the last ice age more frequent among species of animals with large body sizes? A researcher gathers data on the average body mass (in kilograms) of all species known to have existed at that time. These measurements are values of

(a) a categorical variable.

(b) a quantitative variable.

(c) an invalid variable.

(d) a margin of error.

4. An example of a categorical variable is

(a) the name of the college a student attends.

(b) a student's weight in kilograms.

(c) a student's class rank, such as 25th out of 364.

(d) a student's sex (male or female).

(e) Both (a) and (d).

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:



**The next five questions** are related to this situation.

5. The median return during this period is

(a) 5.5 (b) 20.07 (c) 23.0 (d) 25.8 (e) 28.6

6. The third quartile of these returns is

(a) 7.6 (b) 30.5 (c) 31.1 (d) 31.7 (e) 33.4

7. The mean return is

(a) 20.07 (b) 20.69 (c) 22.3 (d) 25.8 (e) 33.4

8. The standard deviation of the returns is

(a) 13.75 (b) 13.98 (c) 14.74 (d) 20.07 (e) 25.8

9. You have similar data on returns on common stocks for all years since 1945. To show clearly how returns have changed over time, your best choice of graph is

(a) a bar graph (b) a line graph (c) a pie chart

(d) a histogram (e) a scatterplot

18. A well-drawn histogram should have

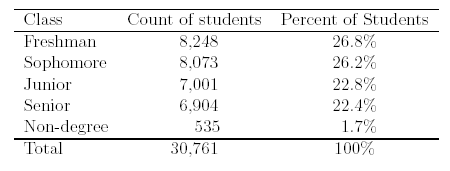
(a) bars all the same width

(b) no space between bars (unless a class has no observations)

(c) a clearly marked vertical scale

(d) all of these

Here is a table of the undergraduate enrollment at a large state university, broken down by class:



21. To make a correct graph of the distribution of students by class, you could use

(a) a bar graph.

(b) a pie chart.

(c) a histogram.

(d) all of (a), (b), and (c).

(e) (a) or (b), but not (c).

42. For a distribution that is skewed to the right, usually

(a) the mean will be larger than the median

(b) the median will be larger than the mean

(c) the first quartile will be larger than the third quartile

(d) the standard deviation will be negative

(e) the minimum will be larger than the maximum

43. Here is a set of data: 1300, 18, 25, 19, *-*7, 24. Which observation is the outlier?

(a) 1300

(b) 25

(c) 19

(d) *-*7

(e) 24

45. You have calculated that the value of the standard deviation for a set of data is -2.5. You can conclude that

(a) the data are skewed to the left

(b) the data are skewed to the right

(c) there is very little variation in the data

(d) all of the observations in the data set are negative

(e) you made an error in your calculations

55. You calculate that the standard deviation of a set of observations is 0. This tells you that

(a) you made an arithmetic mistake.

(b) all the observations have the same value.

(c) there is no straight-line association.

(d) the mean must also be 0.

(e) Both (b) and (d) are true.

57. You are giving a talk about the decay of the American family. You want to show by a graph how the number of divorces has increased between 1970 and 2000. What type of graph should you draw?

(a) A bar graph.

(b) A boxplot.

(c) A histogram.

(d) A line graph.

(e) A pie chart.

69. The five-number summary of a distribution consists of

(a) mean, median, standard deviation, and two quartiles.

(b) minimum, maximum, mean, median, and standard deviation.

(c) minimum, maximum, median, and two quartiles

(d) mean, standard deviation, correlation, and two quartiles

72. The standard deviation is a measure of

(a) the center of a distribution

(b) the variability of a distribution

(c) the association between two variables

(d) the standardized value of a variable

100. The box in the center of a boxplot marks

(a) the full range covered by the data

(b) the range covered by the middle half of the data

(c) the range covered by the middle three-quarters of the data

(d) the span one standard deviation on each side of the mean

(e) the span two standard deviations on each side of the mean

***CHAPTER 5***

Suppose that the BAC of male students at a particular college who drink 5 beers varies from student to student according to a normal distribution with mean 0.08 and standard deviation 0.01. **The next three questions** use this information.

1. The middle 95% of students who drink 5 beers have BAC between

(a) 0.07 and 0.09 (b) 0.06 and 0.10 (c) 0.05 and 0.11 (d) 0.04 and 0.12

2. What percent of students who drink 5 beers have BAC above 0.08 (the legal limit for driving in most states)?

(a) 2.5% (b) 5% (c) 16% (d) 32% (e) 50%

3. What percent of students who drink 5 beers have BAC above 0.10 (the legal limit for driving other states)?

(a) 2.5% (b) 5% (c) 16% (d) 32% (e) 50%

4. SAT scores are normally distributed with mean 500 and standard deviation 100. Julie scores 650. Her standard score is

(a) 150 (b) 15 (c) 1.5 (d) 0.15

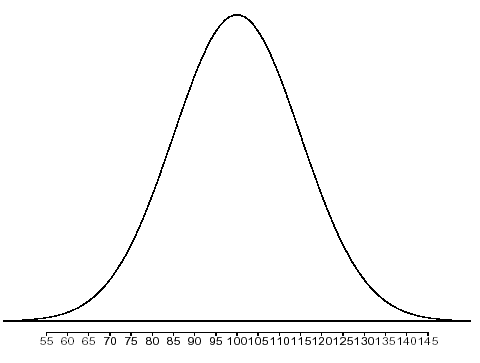
11. The mean of any density curve is

(a) the point where the curvature of the curve changes.

(b) the point at which the curve reaches its highest value.

(c) the point at which the curve would balance if made of solid material.

(d) the point with half the area under the curve to its left and to its right.



13. The mean of the normal curve above is

(a) 80 (b) 90 (c) 100 (d) 110 (e) 120

14. The standard deviation of the normal curve above is

(a) 5 (b) 10 (c) 15 (d) 20 (e) 25

15. If you know the mean and standard deviation of a distribution, do you know the complete shape of the distribution?

(a) Yes, always.

(b) Yes if the distribution is normal, but not in general.

(c) Yes if the distribution is symmetric, but not in general.

(d) No, never.

16. For a normal distribution with mean 20 and standard deviation 5, approximately what percent of the observations will be between 5 and 35?

(a) 50% (b) 68% (c) 95% (d) 99.7% (e) 100%

17. For a normal distribution with mean 20 and standard deviation 5, approximately what percent of the observations will be less than 20?

(a) 50% (b) 68% (c) 95% (d) 99.7% (e) 100%

18. For a normal distribution with mean 20 and standard deviation 5, approximately what percent of the observations will be less than 10?

(a) 99.7% (b) 97.5% (c) 2.5% (d) 95% (e) 99%

28. The distribution of heights of adult men is approximately normal. A man is at the 61.79th percentile. What percent of all men are taller than he is?

(a) 61.79% (b) 38.21% (c) 0.3%

(d) We can't tell from the given information.

(e) We can't tell without the normal table from the text.

29. Which of the following is **least** likely to have a nearly normal distribution?

(a) Heights of all female students taking STAT 001 at State Tech.

(b) IQ scores of all students taking STAT 001 at State Tech.

(c) The SAT Math scores of all students taking STAT 001 at State Tech.

(d) Family incomes of all students taking STAT 001 at State Tech.

(e) Time from conception to birth of all students taking STAT 001 at State Tech.

39. If your score on a test is at the 60th percentile, you know that your score lies

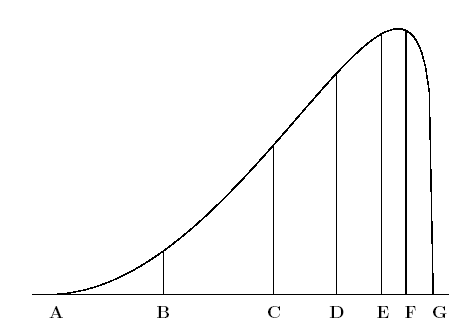
(a) below the first quartile.

(b) between the first quartile and the median.

(c) between the median and the third quartile.

(d) above the third quartile.

(e) Can't say where it lies relative to the quartiles.



45. The figure above is the density curve of a distribution. This distribution is

(a) Roughly symmetric.

(b) Skewed to the left.

(c) Skewed to the right.

(d) Positively correlated.

(e) Negatively correlated.

46. Five of the seven points marked on this density curve make up the five-number summary for this distribution. Which two points are *not* part of the five-number summary?

(a) B and E (b) C and F (c) C and E (d) B and F (e) A and G

***CHAPTER 6***

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:

**The next three questions** are related to this situation.

1. The correlation of U.S. stock returns with overseas stock returns during these years was *r* = 0*.*44. This tells you that

(a) when U.S. stocks rose, overseas stocks also tended to rise, but the connection was not very strong

(b) when U.S. stocks rose, overseas stocks rose by almost exactly the same amount

(c) when U.S. stocks rose, overseas stocks tended to fall, but the connection was not very strong

(d) there is almost no relationship between changes in U.S. stocks and changes in overseas stocks

(e) nothing, because this is not a possible value of *r*

2. If *x* is the return on U.S. stocks and *y* is the return on overseas stocks in the same year, the least-squares regression line for predicting *y* from *x* is *y* = *-*2*.*7 + 0.47*x*. You think U.S. stocks will have a return of 10% in 1999. Using this regression line, you predict that the return on overseas stocks will be

(a) 7.4% (b) *-*2*.*23% (c) 2% (d) 3.17%

3. Stock returns are measured in percent. What are the units of the mean, the median, the quartiles, the standard deviation, and the correlation between U.S. and overseas returns?

(a) all are measured in percent.

(b) all are measured in percent except the standard deviation, which is measured in squared percent.

(c) all are measured in percent except the correlation, which is a number that has no units.

(d) all are measured in percent except the correlation, which is measured in squared percent.

5. Suppose that the correlation between the scores of students on Exam 1 and Exam 2 in a statistics class is *r* = 0*.*7. One way to interpret *r* is to say what percent of the change in Exam 2 scores can be explained by the change in Exam 1 scores. This percent is about

(a) 84% (b) 70% (c) 49% (d) 30%

7. What can we say about the relationship between a correlation *r* and the slope *b* of the least-squares line for the same set of data?

(a) *r* is always larger than *b*

(b) *r* and *b* always have the same sign (+ or *-*)

(c) *b* is always larger than *r*

(d) *b* and *r* are measured in the same units

13. Which statistical measure is **not** strongly affected by a few outliers in the data?

(a) the mean

(b) the median

(c) the standard deviation

(d) the correlation coefficient

16. The least-squares regression line for predicting the percent of a country's females who are illiterate from the percent of males who are illiterate is

female % = 3*.*34 + 1*.*39 male %

In China, 10.1% of men are illiterate. Predict the percent of illiterate women in China.

(a) 4.7% (b) 14% (c) 17.4% (d) 47.8%

17. The equation of the regression line tells us that (on the average) when the male illiteracy rate goes up by 1%, the female rate goes up by

(a) 4.73% (b) 3.34% (c) 1.95% (d) 1.39%

19. You are planning an experiment to study the effect of gasoline brand and vehicle weight on the gas mileage (miles per gallon) of sport utility vehicles. In this study,

(a) gas mileage is a response variable.

(b) gas mileage is an explanatory variable.

(c) gas mileage is a lurking variable.

(d) gas mileage is a categorical variable.

21. A study of 3,617 adults found that those who attend religious services live longer (on the average) than those who don't. Is this good evidence that attending services *causes* longer life?

(a) Yes, because the study is an experiment.

(b) No, because religious people may differ from non-religious people in other ways, such as smoking and drinking, that affect life span.

(c) Yes, because the sample is so large that the margin of error will be quite small.

(d) No, because we can't generalize from 3,617 people to the millions of adults in the country.

22. Which of these is *not* true of the correlation *r* between the lengths in inches and weights in pounds of a sample of brook trout?

(a) *r* must take a value between *-*1 and 1.

(b) *r* is measured in inches.

(c) if longer trout tend to also be heavier, than *r >* 0.

(d) *r* would not change if we measured these trout in centimeters instead of inches.

(e) Both (b) and (d).

23. A correlation cannot have the value

(a) 0.4 (b) *-*0*.*75 (c) 1.5 (d) 0.0 (e) 0.99

24. Which correlation indicates a strong positive straight line relationship?

(a) 0.4 (b) *-*0*.*75 (c) 1.5 (d) 0.0 (e) 0.99

25. A study found that SAT verbal scores were positively associated with first-year grade point averages for liberal arts majors. We can conclude from this that

(a) students who scored high on the SAT verbal test tended to get lower GPAs than those who scored lower on the SAT verbal test

(b) students who scored high on the SAT verbal test tended to get higher GPAs than those who scored lower on the SAT verbal test

(c) we can use the SAT verbal score to accurately predict GPAs for liberal arts majors

(d) grade point averages are higher for older students

(e) the correlation between the SAT verbal score and GPA is higher than 0.5

30. If the least squares regression line for predicting *y* from *x* is *y* = 500 - 20*x*, what is the predicted value of *y* when *x* = 10?

(a) 300 (b) 500 (c) 200 (d) 700 (e) 20

31. Suppose that the least squares regression line for predicting *y* from *x* is *y* = 100 + 1*.*3*x*. Which of the following is a possible value for the correlation between *y* and *x*?

(a) 1.3 (b) *-*1*.*3 (c) 0 (d) *-*0*.*5 (e) 0.5

28. The correlation between two variables is of *-*0.8. We can conclude

(a) one causes the other

(b) there is a strong positive association between the two variables

(c) there is a strong negative association between the two variables

(d) all of the relationship between the two variables can be explained by a straight line

(e) there are no outliers

38. Perfect correlation means all of the following **except**

(a) *r* = *-*1 or *r* = +1.

(b) all points on the scatterplot lie on a straight line.

(c) all variation in one variable is explained by variation in the other variable.

(d) there is a causal relationship between the variables.

(e) each variable is a perfect predictor of the other.

***CHAPTER 7 & 8***

1. In government data, a household consists of all occupants of a dwelling unit. Choose an American household at random and count the number of people it contains. Here is the assignment of probabilities for your outcome:



The probability of finding 3 people in a household is the same as the probability of finding 4 people. These probabilities are marked ??? in the table of the distribution. The probability that a household contains 3 people must be

(a) 0.68 (b) 0.32 (c) 0.16 (d) 0.08

(e) between 0 and 1, and we can say no more.

2. Which of the following statements about a table of random digits is **true**?

(a) If each line contains 40 digits, there will be exactly 4 zeros in every line.

(b) The probability that there are exactly 4 zeros in a line of 40 digits is exactly 0.5.

(c) The number of zeros in a line of 40 digits will vary, but on the average there will be 4 zeros per line.

(d) There can never be 4 zeros in a row because that pattern isn't random.

(e) Both (c) and (d) are true.

3. A friend rolls cheap dice many times. He reports that the probabilities of the possible outcomes are about as follows:



Is this a legitimate probability model?

(a) Yes.

(b) No -- the faces must all have the same probability.

(c) No -- the 3 and 4 faces are opposite each other, so they must have the same probability.

(d) No -- the total probability for all faces is wrong.

(e) No -- not all the values given are possible values for a probability.

Choose an American household at random and ask how many cars and trucks that household owns. Here are the probabilities as of 1997:



4. This is a legitimate assignment of probabilities because it satisfies these rules:

(a) all the probabilities are between 0 and 1.

(b) all the probabilities are between *-*1 and 1.

(c) the sum of all the probabilities is exactly 1.

(d) Both (a) and (c).

(e) Both (b) and (c).

5. What is the probability that a randomly chosen household owns more than one motor vehicle?

(a) 0.96 (b) 0.71 (c) 0.26 (d) 0.25

9. If I toss a fair coin five times and the outcomes are TTTTT, then the probability that tails appears on the next toss is

(a) 0.5

(b) less than 0.5

(c) greater than 0.5

(d) 0

(e) 1

10. If a coin has 0.6 probability coming up tails, the probability that it comes up heads is

(a) 0.5 (b) *-*0.2 (c) 0.4 (d) 0.6 (e) 1.0

12. The probability that the sum is 7 when you roll two dice is 1/6; the probability that the sum is 11 is 1/18. Suppose you play a game where you win if the sum is 7 or 11. What is the probability that you win?

(a) 2/6 (b) 2/18 (c) 7/6 (d) 2/9 (e) 2/24

13. If I toss a fair coin 5000 times

(a) the number of heads will be close to 2500

(b) the proportion of heads will be close to 0.5

(c) the price of oranges will increase

(d) the proportion of heads in these tosses is a parameter

(e) the proportion of heads will be close to 50

20. A household is a group of people living together at the same address. Choose one American household at random and record how many people it contains. Here are the probabilities:



What is the probability that the household chosen contains only one person?

(a) 0.15 (b) 0.25 (c) 0.35 (d) 0.75

(e) Can't tell from the information given.

21. What is the probability that a randomly chosen household contains 4 or more people?

(a) 0.10 (b) 0.16 (c) 0.26 (d) 0.90

(e) Can't tell from the information given.

22. A poker player is dealt poor hands for several hours. He decides to bet heavily on the last hand of the evening on the grounds that after many bad hands he is due for a winner.

(a) He's right, because the winnings have to average out.

(b) He's wrong, because successive deals are independent of each other.

(c) He's right, because successive deals are independent of each other.

(d) He's wrong, because his expected winnings are $0 and he's below that now.

6. In government data, a family consists of two or more persons who live together and are related by blood or marriage. Choose an American family at random and count the number of people it contains. Here is the assignment of probabilities for your outcome:



What is the probability that the family you choose has more than 2 people?

(a) 0.35 (b) 0.42 (c) 0.58 (d) 1.00

(e) Between 0 and 1, and we can say no more.

7. Using the probabilities in the previous question, what is the expected size of the family you draw?

(a) 2 people

(b) 3 people

(c) 3.14 people

(d) 3.5 people

(e) 4.5 people

10. A gambler who keeps placing $1 bets on roulette will, after a very large number of bets, find that his average winnings per bet are close to $0.947. (The house keeps the other $0.053 per bet.) The statistical term for the number $0.947 is

(a) the probability of winning a bet.

(b) the bias of a bet.

(c) a random number.

(d) the expected value of a bet.

12. You play a game with two possible outcomes. Outcome A has probability 0.4 and outcome B has probability 0.6. When B occurs you win $2.00; otherwise, you lose $1.00. What is your expected value for this game?

(a) $2.00 (b) *-*$0.10 (c) $0.20 (d) *-*$0.80 (e) $0.80

***CHAPTER 9***

A recent Gallup Poll asked "Do you consider the amount of federal income tax you have to pay as too high, about right, or too low?" 69% of the sample answered "Too high." Gallup says that for results based on the sample of national adults (*n* = 1*,*055) surveyed April 6-7, 1999, the margin of sampling error is 3 percentage points. **The next two questions** concern this poll.

1. The poll was carried out by telephone, so people without phones are always excluded from the sample. Any errors in the final result due to excluding people without phones

(a) are included in the announced margin of error

(b) are in addition to the announced margin of error

(c) can be ignored, because these people are not part of the population

(d) can be ignored, because this is a nonsampling error

2. If Gallup had used an SRS of size *n* = 1055 and obtained the sample proportion = 0*.*69, you can calculate that the margin of error for 95% confidence would be

(a) 0*.*02 percentage points

(b) 0*.*04 percentage points

(c) 1*.*4 percentage points

(d) 2*.*8 percentage points

(e) 3*.*0 percentage points

The student newspaper at a college asks an SRS of 250 undergraduates, "Do you favor eliminating the carnival from the term-end celebration?" In all 150 of the 250 are in favor. **The next five questions** concern this sample survey.

3. The \_\_\_\_\_\_ you want to estimate is the proportion *p* of all undergraduates who favor eliminating the carnival. That \_\_\_\_\_\_ should read

(a) bias (b) confidence level (c) mean

(d) parameter (e) statistic

4. To estimate *p*, you will use the proportion = 150/250 of your sample who favored eliminating the carnival. The number is a

(a) bias (b) confidence level (c) mean

(d) parameter (e) statistic

5. A 95% confidence interval for the population proportion *p* is

(a) 150 0*.*03 (b) 0*.*6 0*.*03 (c) 150 0*.*06

(d) 0*.*6 0*.*06 (e) 1*.*67  0*.*03

6. A 90% confidence interval based on this same sample would have

(a) the same center and a larger margin of error

(b) the same center and a smaller margin of error

(c) a larger margin of error and probably a different center

(d) a smaller margin of error and probably a different center

(e) the same center, but the margin of error changes randomly

8. You want to estimate the proportion of undergraduates at a college who favor eliminating evening exams. You will choose an SRS. If you enlarge your SRS from 250 to 1000 students, the sample proportion 

(a) will have the same mean and the same standard deviation.

(b) will have smaller bias and the standard deviation will be 1/4 as large.

(c) will have smaller bias and the standard deviation will be 1/2 as large.

(d) will have the same mean and the standard deviation will be 1/4 as large.

(e) will have the same mean and the standard deviation will be 1/2 as large.

9. The phrase "95% confidence" in a Gallup Poll press release means that

(a) our results are true for 95% of the population of all adults.

(b) 95% of the population falls within the margin of error we announce.

(c) the probability is 0.95 that a randomly chosen adult falls in the margin of error we announce.

(d) we got these results using a method that gives correct answers in 95% of all samples.

10. A recent Gallup Poll interviewed a random sample of 1523 adults. Of these, 868 bought a lottery ticket in the past year. A 95% confidence interval for the proportion of all adults who bought a lottery ticket in the past year is (assume Gallup used an SRS)

(a) 0*.*57 0*.*00016

(b) 0*.*57 0*.*00032

(c) 0*.*57 0*.*013

(d) 0*.*57 0*.*025

(e) 0*.*57 0*.*03

16. If the value of the test statistic *z* = 2.5,

(a) conclude that the null hypothesis and the alternative hypothesis are the same

(b) we reject the null hypothesis at the 5% significance level

(c) we fail to reject the null hypothesis at the 5% significance level

(d) we reject the alternative hypothesis at the 5% significance level

(e) should use a different null hypothesis

17. The null hypothesis is

(a) another name for the alternative hypothesis

(b) true with 95% probability

(c) usually a statement of "no effect" or "no difference"

(d) determined by looking at the data

(e) statistically significant

18. If a significance test gives a *P*-value of 0.50,

(a) the margin of error is 0.50

(b) the null hypothesis is very likely to be true

(c) we do not have good evidence against the null hypothesis

(d) we do have good evidence against the null hypothesis

(e) the effect of interest is practically significant

20. If a significance test gives *P*-value 0.005,

(a) the margin of error is 0.005.

(b) the null hypothesis is very likely to be true.

(c) we do not have good evidence against the null hypothesis.

(d) we do have good evidence against the null hypothesis.

(e) the effect of interest is practically significant.

21. The report of a sample survey of 1,014 adults says, "With 95% confidence, between 9% and 15% of all Americans expect to spend more money on gifts this year than last year." The phrase "95% confidence" means

(a) 95% of all Americans will spend between 9% and 15% of what they spent last year.

(b) 9% to 15% of all Americans will spend 95% of what they spent last year.

(c) there is a 95% chance that the percent who expect to spend more is between 9% and 15%.

(d) the method used to get the interval from 9% to 15%, when used over and over, produces intervals which include the true population percentage 95% of the time.

(e) we can be 95% confident that the method used to get the interval always gives the right answer.

22. A sample survey finds that 30% of a sample of 874 Ohio adults said good health was the thing they were most thankful for. If that sample were an SRS from the population of all Ohio adults, what would be the 99% confidence interval for the percent of all Ohio adults who feel that way?

(a) 25% to 35% (b) 26% to 34% (c) 27% to 33%

(d) 28% to 32% (e) 29% to 31%

23. If the 874 people in the previous question had called a 900 number to give their opinions, how would this affect your answer?

(a) Not at all, because the width of the confidence interval depends only on the sample size, and not on the population size.

(b) Not at all, because the width of the confidence interval depends only on the sample size, and not on how the sample was obtained.

(c) It would be wider because voluntary response polls have a bigger margin of error than SRSs.

(d) It would be narrower because voluntary response polls are less variable than SRSs.

(e) A confidence interval makes no sense for a voluntary response sample.

24. The name for the pattern of values that a statistic takes when we sample repeatedly from the same population is

(a) the bias of the statistic.

(b) the sampling distribution of the statistic.

(c) the scale of measurement of the statistic.

(d) the variability of the statistic.

(e) the sampling error.

28. A CBS News/New York Times opinion poll asked 1,190 adults whether they would prefer balancing the Federal budget over cutting taxes; 702 of those asked said "Yes." Take the sample to be an SRS from the population of all adults. Which of these is a correct 95% confidence interval for the proportion of all adults who prefer balancing the budget over cutting taxes?

(a) 0*.*59 0*.*0004

(b) 0*.5*9 0*.*014

(c) 0*.5*9 0*.*0186

(d) 0*.5*9 0*.*0285

(e) 0*.5*9 0*.*037

29. Suppose that in fact 62% of all adults favor balancing the budget over cutting taxes. The number 62% is

(a) a bias.

(b) a margin of error.

(c) a statistic.

(d) a parameter.

(e) a coefficient of variation.

30. Suppose that in fact 62% of all adults favor balancing the budget over cutting taxes. If you take a large number of SRSs of size 1,190, the sample proportions who favor balancing the budget will vary. Some will be lower than 62% and some will be higher, but the average sample result will be very close to 62%. This fact is called

(a) small bias.

(b) small margin of error.

(c) high variability.

(d) large bias.

(e) low variability.

***CHAPTER 10***

6. A Census Bureau report on the income of Americans says that with 95% confidence the median income of all U.S. households in 1997 was $37,005 with a margin of error of **$342.

This means that

(a) 95% of all households had incomes in the range $37*,*005** $342.

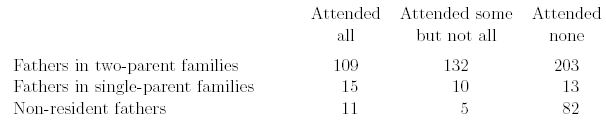
(b) we can be sure that the median income for all households in the country lies in the range $37,005 ** $342.

(c) 95% of the households in the sample interviewed by the Census Bureau had incomes in the range $37,005 ** $342.

(d) the Census Bureau got the result $37,005 ** $342 using a method that will cover the true median income 95% of the time when used repeatedly.

A study of fathers' involvement in their children's education interviews a sample of fathers of school-age children. One question concerns regularly scheduled parent-teacher conferences.

Here is a two-way table of the results:



**The next five questions** concern these data.

17. How many fathers were interviewed?

(a) 135 (b) 444 (c) 580 (d) 444,000 (e) 580,000

18. How many fathers from single-parent families were interviewed?

(a) 38 (b) 444 (c) 580 (d) 38,000 (e) 444,000

19. What percent of single-parent fathers attended all the conferences?

(a) about 39% (b) about 23% (c) about 11% (d) about 2.5%

20. We can summarize the relationship between family status and fathers' attendance at parent- teacher conferences by saying

(a) Few fathers are single parents, more are nonresident, and most are in two-parent families.

(b) Fathers in two-parent families are much more likely to attend than other fathers, and nonresident fathers rarely attend.

(c) Nonresident fathers rarely attend, and single-parent fathers are somewhat more likely to attend than fathers in two-parent families.

(d) Most fathers attended no school conferences, and only about a quarter attended all.

21. The chi-square statistic for this two-way table is ** = 54*.*8 with *P*-value *P <* 0*.*001. We can conclude that

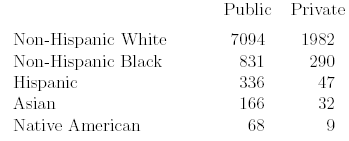
(a) There is very strong evidence that family status and fathers' attendance at school conferences are related in the population of all fathers.

(b) There is very strong evidence that family status and fathers' attendance at school conferences are related among the fathers in this sample.

(c) We lack strong evidence that family status and fathers' attendance at school conferences are related in the population of all fathers.

(d) We lack strong evidence that family status and fathers' attendance at school conferences are related among the fathers in this sample.

Here is a two-way table of the numbers of college students (U.S. citizens only) classified by racial/ethnic group and by whether they attend a public or a private college. The entries are in thousands of students. **The next four questions** concern these data.



22. How many non-Hispanic black college students are there?

(a) 1121 (b) 11,210 (c) 112,100 (d) 1,121,000 (e) 11,210,000

23. What percent of non-Hispanic black students attend public colleges?

(a) 9.8% (b) 10.4% (c) 13.3% (d) 26% (e) 74%

24. There is no reason to use the chi-square test on these data, because

(a) The hypothesis that chi-square tests makes no sense here.

(b) The data describe the entire population of college students, so there is no inference from sample to population.

(c) The sample size is so large that any test will certainly be highly significant.

(d) The sample size is too small to allow use of the chi-square test.

(e) This is a voluntary response sample.

25. What percent of public college students are non-Hispanic blacks?

(a) 9.8% (b) 10.4% (c) 13.3% (d) 26% (e) 74%

26. The margin of error for a 95% confidence interval is 2.8. If we decrease the confidence level

to 90%, the margin of error will be

(a) biased (b) 99%

(c) 2.8 (d) smaller than 2.8

(e) larger than 2.8